

Stable Nash Equilibria in the Gale-Shapley Matching Game

Sushmita Gupta*

Kazuo Iwama*

Shuichi Miyazaki*

Abstract

In this article we study the stable marriage game induced by the men-proposing Gale-Shapley algorithm. Our setting is standard: all the lists are complete and the matching mechanism is the men-proposing Gale-Shapley algorithm. It is well known that in this setting, men cannot cheat, but women can. In fact, Teo, Sethuraman and Tan [TST01], show that there is a polynomial time algorithm to obtain, for a given strategy (the set of all lists) Q and a woman w , the best partner attainable by changing her list. However, what if the resulting matching is not stable with respect to Q ? Obviously, such a matching would be vulnerable to further manipulation, but is not mentioned in [TST01]. In this paper, we consider (safe) manipulation that implies a stable matching in a most general setting. Specifically, our goal is to decide for a given Q , if w can manipulate her list to obtain a strictly better partner with respect to the true strategy P (which may be different from Q), and also the outcome is a stable matching for P .

1 Introduction

Matchings under preferences is an extensively studied area of theoretical and empirical research that has a wide range of applications in economics and social sciences. The most popular and standard problem in this field is the *stable marriage problem* (SMP), introduced by Gale and Shapley [GS62], where we have two parties; a set of men and a set of women. Each man has a preference list that orders women according to his preference, and similarly each woman also has a preference ordering for the men. The goal is to find a stable matching, a matching without any *blocking pair*, where a blocking pair is a pair of a man and a woman who prefer each other to their current partners.

The goal of each player is of course to obtain a good partner (to steal a good partner from colleagues). Thus, the stable marriage is a typical game played by selfish players and its game-theoretic aspects have been studied for a long time (see Roth and Sotomayor [RS90]). For instance, it is well known that if we use the men-proposing Gale-Shapley algorithm (GS-M, in short) as the matching mechanism, then men do not benefit by cheating, *i.e.* any change of a single list on the men's side does not give him a strictly better partner. Also, if we allow individuals to truncate their lists (*i.e.* delete one or more persons from the list), then there are several (relatively easy) ways to manipulate to obtain a better outcome.

For the special case, when all lists are complete, we have known about instances of SMP (see page 65 of [GI89], for example), where a woman can improve her outcome by permuting her true preference list, while all others use their true lists. But little else was known about the property of *permutation strategies*, such as the time complexity of finding such a list. It is due to the work of Teo *et al.* [TST01] that this question was resolved. They proved that there is a poly-time algorithm to obtain, for a given set of complete preference lists for all men and women, and a particular woman w , the best partner that w can attain by using one of

*Kyoto University, Japan.

Emails: sushmita.gupta@gmail.com, iwama@kuis.kyoto-u.ac.jp, shuichi@media.kyoto-u.ac.jp

the $n!$ permutations of her true preference list. As a corollary, it is also possible to decide in polynomial time if the set of true preference lists is a Nash equilibrium. Thus, [TST01] was a nice start towards progress in this topic and their proof is also nontrivial. However, their setting that there is no distinction between the “true” preference list of a player and the actual preference list submitted by him/her, is not standard in game theory. This is supported by theory, by the result that there is no mechanism in which the best strategy for all the agents is to tell the truth [RS90, Thm 4.4]. For GS-M it is in the best interest for the men to tell the truth, but that is not the case for the women. Hence, we have to consider the possibility that true preference is different from stated preference.

Recall that in the general setting of game theory, we have a set of players and their strategies. The outcome of each player i for his/her strategy s_i , is determined by a function of s_i and all the other players’ strategies. In the case of stable marriage, the strategy of each player is his/her preference list, and that can be one of $n!$ different ones and the player’s outcome is determined by final matched partner at the end of the matching algorithm, (GS-M in our case). The true preference list is necessary for assessing an agent’s profit or loss; the actual stated strategy (submitted list) may of course be very different from the true list. In other words, a stated list of a woman w and her true lists are effectively independent; the latter is necessary for computing the quality (*i.e.* benefit, if any) of the outcome for w , and as it turns out, has another important role.

Consider the strategies depicted in Figure 1: We have five men $\{1, \dots, 5\}$, five women $\{a, \dots, e\}$, and their current preference lists, given by Q_1 . Recall that each person has his/her true preference list, and that is depicted by P in the table. We want to see how vulnerable Q_1 is against manipulation. GS-M applied to Q_1 yields $\mu_{Q_1} = \{(1, a), (2, b), (3, c), (4, d), (5, e)\}$, which is stable with respect to P (called P -stable), since $\mu_P = \mu_{Q_1}$. However, it is not stable in the Nash equilibrium sense, since d can improve her outcome by changing her list to $Q_2(d)$, resulting in $\mu_{Q_2} = \{(1, a), (2, b), (3, c), (4, e), (5, d)\}$. Her new partner 5 is better than 4 in terms of $P(d)$ and the matching is again P -stable. It is noteworthy that μ_{Q_2} is the women-optimal P -stable matching, hence it admits no P -stable manipulation strategies.

But then again, another woman a gets into the game, against Q_2 , and using $Q_3(a)$. Her new partner is 2, who is ranked first in her true list. Unfortunately, though, μ_{Q_3} is not P -stable, with blocking pairs $(2, b)$ and $(3, b)$. This leaves open the possibility of manipulation by b , given by $Q_4(b)$, whereby b is matched to 2, and a to 3. Thus, a actually has a lower payoff in Q_4 than Q_2 , thus rendering her manipulating of Q_3 counter-productive. This is why we posit that manipulation strategies that are P -stable add a layer of security for the manipulation, since unstable manipulation will always leave open the possibility of others to manipulate, specifically those women who are in a P -blocking pair.

Our Contributions: Motivated by these strategic considerations, we introduce a *stable Nash equilibrium*, which is a relaxation of a Nash equilibrium, but nonetheless sufficiently limits an agent’s incentive to manipulate the market. Suppose that the stable matching game is induced by the men-optimal stable matching mechanism. For a given strategy Q and a true strategy P , Q is said to be a *P -stable Nash equilibrium* (in this game) if and only if there is no woman w such that (i) w can manipulate her list, (while all others retain their Q -lists) and obtain a better partner with respect to $P(w)$, and (ii) the resulting matching is P -stable. Note that there are several cases that people may try to improve their outcome but actually fail under (i) and (ii). For instance, (a) w manipulates her list, and indeed gets matched to x , who is better with respect to Q , but not with respect to P , and (b) w manipulates her list, gets indeed a better partner x with respect to P , but the resulting matching is not P -stable, for instance Q_3 , in our example.

Our main result in this paper is the existence of a polynomial time algorithm that determines

Men	$P(M)$	Women	$P(W)$	$Q_1(W)$
1 :	$a \ b \ d \ e \ c$	$a :$	$2 \ \underline{1} \ 4 \ 5 \ 3$	$2 \ \underline{1} \ 4 \ 5 \ 3$
2 :	$b \ a \ d \ e \ c$	$b :$	$\underline{2} \ 3 \ 1 \ 4 \ 5$	$1 \ \underline{2} \ 4 \ 5 \ 3$
3 :	$a \ b \ d \ e \ c$	$c :$	$1 \ 2 \ 4 \ 5 \ \underline{3}$	$1 \ 2 \ 4 \ 5 \ \underline{3}$
4 :	$d \ e \ a \ b \ c$	$d :$	$5 \ \underline{4} \ 1 \ 2 \ 3$	$5 \ \underline{4} \ 1 \ 2 \ 3$
5 :	$e \ d \ a \ b \ c$	$e :$	$4 \ \underline{5} \ 1 \ 2 \ 3$	$4 \ \underline{5} \ 1 \ 2 \ 3$

True strategy $P = (P(M), P(W))$ and stated strategy $Q_1 = (P(M), Q_1(W))$. The M -optimal matching $\mu_P = \mu_{Q_1} = \{(1, a), (2, b), (3, c), (4, d), (5, e)\}$, thus, μ_{Q_1} is P -stable.

Men	$P(M)$	Women	$Q_2(W)$	$Q_3(W)$	$Q_4(W)$
1 :	$a \ b \ d \ e \ c$	$a :$	$2 \ \underline{1} \ 4 \ 5 \ 3$	$\underline{2} \ 3 \ 1 \ 4 \ 5$	$2 \ \underline{3} \ 1 \ 4 \ 5$
2 :	$b \ a \ d \ e \ c$	$b :$	$1 \ \underline{2} \ 4 \ 5 \ 3$	$\underline{1} \ 2 \ 4 \ 5 \ 3$	$\underline{2} \ 3 \ 1 \ 4 \ 5$
3 :	$a \ b \ d \ e \ c$	$c :$	$1 \ 2 \ 4 \ 5 \ \underline{3}$	$1 \ 2 \ 4 \ 5 \ \underline{3}$	$\underline{1} \ 2 \ 4 \ 5 \ 3$
4 :	$d \ e \ a \ b \ c$	$d :$	$\underline{5} \ 3 \ 4 \ 1 \ 2$	$\underline{5} \ 3 \ 4 \ 1 \ 2$	$5 \ 3 \ \underline{4} \ 1 \ 2$
5 :	$e \ d \ a \ b \ c$	$e :$	$\underline{4} \ 5 \ 1 \ 2 \ 3$	$\underline{4} \ 5 \ 1 \ 2 \ 3$	$4 \ \underline{5} \ 1 \ 2 \ 3$

The P -blocking pairs in μ_{Q_3} are $(2, b)$ and $(3, b)$. Manipulation by b gives $Q_4 = (Q_3(-b), P(b))$, resulting in $\mu_{Q_4} = \{(1, c), (2, b), (3, a), (4, d), (5, e)\}$.

Figure 1: Manipulation by d , followed by a , and then b .

if Q is a P -stable Nash equilibrium. Our algorithm has two stages. In the first stage we obtain triples $(w, x, Q(x; w))$, where w is a woman, $x (\neq \mu_Q[w])$ is a man, and $Q(x; w)$ is a list of w such that $\mu_{(Q(-w), Q(x; w))}$ matches x to w . Here, the preference lists $(Q(-w), Q(w, x))$, is such that only w 's list is changed to $Q(x; w)$, while each $a \in M \cup W \setminus \{w\}$ uses $Q(a)$ (formal definitions in Section 2). It should be noted that there may be many different lists of w that if used instead, would also match w to x , but this algorithm outputs just one of them. The algorithm itself is basically the same as [TST01], but the analysis is completely new. Recall that [TST01] only guarantees that it outputs a similar triple $(w, x, Q(x; w))$, where x is the *best* possible partner of w (with respect to $Q(w)$). Our arguments are completely new, proving the more general statement, that *all* partners attainable by GS-M can be found using the algorithm.

The second stage is to decide, for each triple $(w, x, Q(x; w))$ obtained in the first stage, whether or not there exists a list for w , say $Q'(x; w)$, which if used instead of $Q(w)$ would result in w to x , with the condition that $\mu_{(Q(-w), Q'(x; w))}$ is P -stable. List $Q(x; w)$ is an obvious candidate for such a $Q'(x; w)$, but there may be (exponentially) many others, and it is possible that $Q'(x; w)$ results in a P -stable matching but $Q(x; w)$ does not. For this, we prove a key lemma that is interesting in its own right. Specifically, we show that $\mu_{(Q(-w), Q'(x; w))}$ is unique regardless of the list $Q'(x; w)$. Therefore, to detect if Q is P -stable Nash equilibrium, it suffices to check if $\mu_{(Q(-w), Q(x; w))}$ is P -stable for all the (at most n^2) triples obtained in the first stage.

Related Work: We denote the true preferences of (M, W) by P , and for an individual agent $a \in M \cup W$, the preference list of a by $P(a)$. The M -optimal and W -optimal matching **w.r.t** a given strategy Q is denoted by μ_Q .

For a stated strategy Q , Dubins and Freedman [DF81], proved that there is no coalition C of men, who have a manipulation strategy $P' = (P(-C), P'(C))$ so that the outcome P' -stable, and is strictly better than μ_P , in $P(m)$, for each $m \in C$. Demange, *et al.* [DGS87] extend this result to include women in the coalition C , showing that there is no P' -stable matching μ' such that every agent $a \in C$, prefers $\mu'[a]$, (in $P(a)$) to his/her partner in *every* P -stable matching.

For truncation strategies it was shown by Gale and Sotomayor [GS85] that if there are at least two P -stable matchings, then there is a woman w who has a unilateral manipulation strategy $Q' \in \mathcal{Q}_w$ that gives a strictly better outcome than μ_P , in $P(w)$. If $C = W$, then there

is a truncation strategy $P' = (P(M), P'(W))$, such that $\mu_{P'}$, is the women-optimal matching for P . Considerable work on truncation strategies have been undertaken, see [Ehl08, RR99] for motivations and applications. In fact, up until the late 1980s, analyses of manipulation strategies of women centred almost exclusively around truncation strategies.

As stated earlier, (to the best of our knowledge) Teo *et al.* [TST01] was the first to consider permutations of a agent's true lists as a type of manipulation strategy. They gave a polynomial time algorithm that for a given $w \in W$, computes the best manipulated partner (in $P(w)$) that w can attain using GS-M on any strategy $(P(-w), P'(w)) \in \mathcal{P}_w$.

Immorlica and Mahidian [IM05], allowing the men's preference lists to have ties but bounded by a constant, and drawn from an arbitrary probability distribution, while the women's lists are arbitrary and complete; show that with high probability, truthfulness is the best strategy for an any agent, assuming everybody is being truthful as well.

Kobayashi and Matsui [KM10] consider the possibility that a coalition C of women have a manipulation strategy $P' = (P(M), P'(W))$, containing complete lists, such that $\mu_{P'}$ yields specific partnerships for the members of C . The situation manifests in two specific forms, depending on the nature of the input. In the first case, the input consists of the complete lists of all men, a partial matching (some agents may be unmatched) μ' , and complete lists of the subset of women who are unmatched in μ' , denoted by $W \setminus C$. The problem is to test whether there exists a permutation strategy for each woman in C , such that for the combined strategy $P' = (P(-C), P'(C))$, $\mu_{P'}$ is a perfect matching that extends μ' . In the second case, the input consists of the lists of all men, a perfect matching μ , and lists for women in $W \setminus C$. The problem is to test if there are permutation strategies for the women in C , such that for $P' = (P(-C), P'(C))$, matching $\mu'_P = \mu$.

Pini *et al.* [PRVW11] show that for an arbitrary instance of the stable marriage problem, there is a stable matching mechanism for which it is NP-hard to find a manipulation strategy. More recently, Deng, *et al.* [DST15], drawing on [KM10] have discussed the possibility of a coalition of women permuting their true lists, while others state theirs truthfully, so as to produce a matching that is Pareto optimal for the members of the coalition. They also give an $O(n^6)$ algorithm to compute a strong Nash equilibrium that is strongly Pareto optimal for all the coalition partners.

Our work deals with manipulation by women, in the game induced by GS-M, where truth-telling is the best strategy for men. For a background on manipulation strategies for men, we point the reader to [Hua06, III⁺13] and [Man13].

2 Preliminaries

We always use M to denote the set of n men $\{m_1, m_2, \dots, m_n\}$ and W the set of n women $\{w_1, w_2, \dots, w_n\}$. Our matching mechanism is always the men-proposing Gale-Shapley algorithm: On the man's side, a man who is *single*, *proposes* to the woman who is at the top of his current list. At the woman's side, when a woman w receives a proposal from a man m , she *accepts* that proposal if she is single. Otherwise, if she prefers m to her current partner m' (m' to m , resp.), then w accepts m (m' , resp.) and *rejects* m' (m , resp.). If m is rejected by w , then m becomes single again and w is removed from the m 's list. The algorithm terminates when there is no single man (for more details, see [GI89]). We assume that if there are two or more single men, the man with the smallest index does a proposal, thus making the procedure deterministic.

A *strategy* Q is a set of *preference lists* (or simply *lists*) of all the men in M and all the women in W . For a person x in $M \cup W$, $Q(x)$ denotes the x 's list in the strategy Q . For a given strategy Q , suppose that *only* w changes her list from $Q(w)$ to $Q'(w)$. We denote the resulting strategy by $Q' = (Q(-w), Q'(w))$, and use \mathcal{Q}_w to denote the family of all such strategy Q' . Note

that all lists considered in this article are complete, *i.e.*, they are permutations of n men or n women.

Let μ be a (perfect) matching between M and W and Q be a strategy. Then we say that μ is Q -stable if there is no Q -blocking pair. Here Q -blocking pair is a pair (m, w) of a man and a woman such that m prefers, in (terms of the preference ordering in) $Q(m)$, w to his partner in μ (denoted by $\mu[m]$) and w prefers, in $Q(w)$, m to $\mu[w]$. If w strictly prefers m_1 to m_2 , in $Q(w)$, then we denote that as $m_1 > m_2$ in $Q(w)$. We use $m_1 \geq m_2$, if $m_1 > m_2$, or $m_1 = m_2$.

For a strategy Q , μ_Q denotes the man-optimal stable matching, computed by the Gale-Shapley algorithm. If a man m proposes to a woman w during this procedure, then we say that m is *active* in $Q(w)$ (formally speaking we should say m is active in $Q(w)$, during the computation of μ_Q , but for the sake of brevity, we will omit strategy Q when it is obvious from the context.)

Recall that a woman w changes her list $Q(w)$ for the purpose of manipulation. For a subset $M' \subseteq M$, let I be an ordering of men in M' . Then, $Q(I; w)$ denotes a permutation of $Q(w)$, where the men in M' are at the front in the order in which they appear in I . An ordered (sub)list such as I , is called a *tuple*, and for any given tuple I , we define $Q(I; w) = [I, Q(w) \setminus I]$. For example, if $Q(w) = (1, 2, 3, 4, 5, 6)$ and $I = (5, 2)$, then $Q(I; w) = (5, 2, 1, 3, 4, 6)$. Now we are ready to introduce our main concept.

We are given a strategy Q and a *true* strategy P . Then for a woman $w \in W$, a strategy $Q' \in \mathcal{Q}_w$ is said to be a *unilateral manipulation strategy* of w , if $\mu_{Q'}[w] > \mu_Q[w]$ in $P(w)$, *i.e.* w strictly prefers the outcome of $\mu_{Q'}$ to μ_Q , with respect to her true preference/strategy. If $\mu_{Q'}$ is a P -stable matching, then Q' is said to be a *P -stable manipulation strategy* of w . A strategy Q is said to be a *P -stable Nash equilibrium* if there does not exist $w \in W$ having a P -stable unilateral manipulation strategy $Q' \in \mathcal{Q}_w$.

In this paper, we consider the following problem:

Problem: Given a true preference P , and a stated preference Q , determine if Q is a P -stable Nash equilibrium.

3 Listing active men

Now our goal is to design an algorithm that, for two given strategies, a stated strategy Q and a true strategy P , answers if Q is a P -stable Nash equilibrium. To do so, we first design an algorithm that outputs the set $N_w(Q)$ of all possible partners m of a given fixed woman w such that there is a (manipulated) strategy $Q' = (Q(-w), Q'(w))$, for which the men-optimal stable matching (*i.e.* men-proposing GS algorithm), will match w to m . By using this algorithm n times, we can obtain $\{N_{w_1}(Q), \dots, N_{w_n}(Q)\}$. The use of this set to prove our main result is explained in the next section.

See Algorithm 1, which is basically the same as the one given by Teo *et al.* [TST01]: Suppose that $Q(w) = (1, 2, 3, 4, 5, 6, 7, 8)$ and the first proposal comes from man 5. Then the algorithm adds 5 to N and call **Explore**($Q(5; w)$), namely it executes men-proposing Gale-Shapley algorithm (GS-M, in short) after moving man 5 to the front of $Q(w)$. In general, procedure **Explore** takes as a parameter, $Q(x, I; w)$, a preference list of w . As per our notation, x is at the front of this list, followed by the sublist I and then the rest of the men, thus, defining the strategy $Q' = (Q(-w), Q(x, I; w))$. **Explore**($Q(x, I; w)$) executes GS-M for the strategy Q' and produces the set of men A who propose to w after x . Now for each $y \in A$, we check if y is “new” (*i.e.*, not yet in N). If so, we add y to N and call **Explore** recursively after moving y to the top of $Q(x, I; w)$, else, we do nothing.

Since **Explore** is called only once for each man in N , its time complexity is obviously at most $n \times T(\text{GS})$, where $T(\text{GS})$ is the time complexity of one execution of GS-M, thus, overall

it is $O(n^3)$. The nontrivial part is the correctness of the argument, which we shall prove now.

Theorem 1. *For a strategy Q and a woman $w \in W$, Algorithm 1 produces $N = \{m \in M \mid \exists Q' \in \mathcal{Q}_w, \text{ s.t. } \mu_{Q'}[w] = m\}$ and for each $m \in N$, a list $Q(m, I; w)$ such that for some partial list I , m is active in $Q(I; w)$.*

Proof. Let $Q'(w)$ be an arbitrary permutation of n men and Q' be the strategy $(Q(-w), Q'(w))$. It is enough to prove if a man $x \in M$ proposes to w during the computation of $\mu_{Q'}$ (i.e. x is active in $Q'(w)$), then x is added to N during the execution of Algorithm 1.

Here we need two new definitions: Suppose that x_1, x_2, \dots, x_t is a sequence of proposals received by w during the computation of $\mu_{Q'}$. Then this sequence is called an *active sequence* for $Q'(w)$, denoted by $AS'(w)$. Also suppose that y_1, y_2, \dots, y_s is a maximal subsequence of $AS'(w)$ such that $y_1 = x_1$ and $y_s > y_{s-1} > \dots > y_1$ in $Q'(w)$. Then, this is called the *increasing active subsequence* for $Q'(w)$ and is denoted by $IAS'(w)$. As an example, let $Q'(w) = (1, 2, 3, 4, 5, 6, 7, 8, 9)$ and $AS'(w) = 5, 6, 3, 4, 2, 8$. Then $IAS'(w) = 5, 3, 2$. Now consider a different list $Q''(w) = (1, 2, 3, 5, 9, 8, 4, 6, 7)$, thus, $Q'(w) \neq Q''(w)$. Then, we can observe that the active sequence and the increasing active subsequence for $Q''(w)$ are identical to those of $Q'(w)$. The reason being the following. The lists in Q' and Q'' are the same except that of w 's, so the first proposal for w must come from the same man regardless of w 's list. Since 5 is accepted by w (both lists are complete), the next proposal should also be from the same man, 6. Now since $5 > 6$ in both $Q'(w)$ and $Q''(w)$, 6 is rejected in both $Q'(w)$ and $Q''(w)$ and thus, the next proposal must also be same, and so on, we continue. This observation leads us to the following lemma.

Lemma 1. *For strategies $Q', Q'' \in \mathcal{Q}_w$, let x_1, x_2, \dots, x_p and u_1, u_2, \dots, u_q , denote the active sequences for $Q'(w)$ and $Q''(w)$ respectively and let y_1, y_2, \dots, y_s and v_1, v_2, \dots, v_t denote the corresponding increasing active subsequence. Then, the following conditions must hold.*

- (a) $x_1 = y_1 = u_1 = v_1$.
- (b) *For an arbitrary l ($l \leq p$ and $l \leq q$), we consider the prefixes of the active sequences up to position l and the prefixes of the corresponding increasing active subsequences, denoted by y_1, \dots, y_l and v_1, \dots, v_l . Then, if $x_i = u_i$ for all $i \leq l$ and $y_k = v_k$ for all $k \leq l$, then $x_{l+1} = u_{l+1}$.*

Proof. By definition, $x_1 = y_1$ and $u_1 = v_1$. Recall that all the lists in Q' and Q'' are the same except those for w . Furthermore, we use a fixed tie-breaking protocol in the deterministic GS algorithm. Hence, $x_1 = u_1$, follows directly.

To prove condition (b), let $y_2 = x_{i_1+1}, y_3 = x_{i_2+1}, \dots$, and so on. Then we can write $AS'(w)$ as follows, where $x_2, \dots, x_{i_1}, x_{i_1+2}, \dots, x_{i_2}, \dots$ may be empty.

$$AS'(w) = y_1, x_2, \dots, x_{i_1}, y_2, x_{i_1+2}, \dots, x_{i_2}, \dots, y_j, x_{i_{j-1}+2}, \dots, x_l, x_{l+1}, \dots$$

Now one can see that y_1 is accepted, all of x_2, \dots, x_{i_1} are rejected since they are after y_1 in the list by definition. This continues as y_2 is accepted, $x_{i_1+2}, \dots, x_{i_2}$ rejected, and so on. Now, consider $AS''(w)$, depicted below.

$$AS''(w) = v_1, u_2, \dots, u_{i_1}, v_2, u_{i_1+2}, \dots, u_{i_2}, \dots, v_j, u_{i_{j-1}+2}, \dots, u_l, u_{l+1}, \dots$$

By the assumption, these two sequences are identical up to position l , so acceptance or rejection for each proposal follows identically, as discussed above. Therefore, the configuration (see below) of GS-M for Q' at the moment when x_l proposes to w and the configuration for Q'' when u_l proposes to w are exactly the same. A configuration consists of (i) the lists of all

men (recall that some entries are removed when proposals are rejected), (ii) the set of single men, and (iii) the current temporal partner of each woman. (Formally this should be shown by induction, but it is straight forward and may be omitted). Also the acceptance/rejection for x_l and u_l is the same. Thus in either case, the execution of the (deterministic) GS-M is exactly the same for Q' and Q'' until w receives proposal from x_{l+1} and u_{l+1} , respectively. Hence, x_{l+1} and u_{l+1} , should be equal and the lemma is proved. \square

Now let us look at the execution sequence of Algorithm 1 while comparing it with the execution sequence of GS-M on Q' . We assume that the active sequence and increasingly active sequence for Q' are $AS'(w) = x_1, x_2, \dots, x_p$ and $IAS'(w) = y_1, y_2, \dots, y_s$, respectively. By 1, the first proposal to w is always y_1 , so the algorithm starts with **Explore**($Q(y_1; w)$) (we simply say the algorithm invokes $Q(y_1; w)$), and $N = \{y_1\}$, at the very beginning.

Now we note that it is quite easy to see that the active sequence for $Q(y_1; w)$ should be $y_1, x_2, \dots, x_{i_1}, \dots$, *i.e.* it should be identical to that of $Q'(w)$ up to the position i_1 . The reason is as follows. We already know the first active man is always y_1 and that is also the first symbol in the increasing active sequence of both. Thus we can use 1, to conclude that the second symbols should also be the same. Since, x_2 is not in $IAS'(w)$, meaning that it is rejected, which is also the same in $Q(y_1; w)$ since y_1 is at the top of the sequence. Thus the third symbol is the same in both and so on up to position i_1 . Then the next symbol in $AS'(w)$ is y_2 and its that is also active in $Q(y_1; w)$, meaning $Q(y_2, y_1; w)$ is invoked by the algorithm. (The algorithm also invokes $Q(x_2, y_1; w)$, $Q(x_3, y_1; w), \dots, Q(x_{i_1}, y_1; w)$ also, but these are not important for us at this moment).

We again consider the active sequence for $Q(y_2, y_1; w)$ and by the same argument presented earlier, we can conclude that it is identical to $AS'(w)$ up to position i_2 and so y_3 is found to be an active man. Hence, $Q(y_3, y_2, y_1; w)$ is invoked if y_3 was not already present in N .

Continuing like this, we note that if $Q(y_s, y_{s-1}, \dots, y_1; w)$ is invoked, then we are done since its active sequence is identical to that of $Q'(w)$. However, this case happens only if each y_i , $2 \leq i \leq s$, is a brand new active man found during the invocation of $Q(y_{i-1}, \dots, y_1; w)$. If one them is not new then the subsequent lists are not invoked, and yet, we are assured due to 2 that Algorithm 1 will detect all the active men in $Q'(w)$.

2 is rather surprising and maybe of independent interest. For two lists $Q'(w)$ and $Q''(w)$, that are distinct and arbitrary orderings on men, we assume nothing about the execution of GS-M on the two lists except that a particular man x is active in either list. Yet, we are able to show that a man who proposes to w after x when $Q'(x; w)$ is used, must also propose when $Q''(x; w)$ is used. This result allows us to focus solely on active men that have been discovered in the current invocation of **Explore**, thereby restricting the number of recursion steps to $O(n)$.

Lemma 2. *For strategies Q' and Q'' in \mathcal{Q}_w , suppose that x is active in both $Q'(w)$ and $Q''(w)$. Then a man who is active in $Q'(x; w)$ after x , is also active in $Q''(x; w)$.*

Proof. Let y denote a man who is active in $Q'(x; w)$ after x . Consider the following strategies,

$$Q_y = (Q'(-w), Q'(y, x; w)) \text{ and } Q_x = (Q''(-w), Q''(x; w)),$$

and the M -optimal matchings denoted by μ_1 and μ_2 , respectively; *i.e.* $\mu_1[w] = y$ and $\mu_2[w] = x$.

The lists are complete, and $Q(y) = P(y)$, so either $\mu_2[y] > w$, or $w > \mu_2[y]$ in $P(y)$. If $\mu_2[y] > w$ then matching μ_2 is Q_y -stable. Thus, both μ_1 and μ_2 are Q_y -stable, with μ_1 being the men-optimal matching. But, then $w = \mu_1[y] \geq \mu_2[y]$ in $P(y)$, contradicting the initial assumption that $\mu_2[y] > w$. Therefore, we conclude that $w > \mu_2[y]$ in $P(y)$, implying that y should propose to w before $\mu_2[y]$ or that y is active in $Q''(x; w)$. Hence, the proof is complete. \square

The next lemma will complete the proof. Note that we are still using $AS'(w)$, as given in the proof of 1, and using that to define the following lists for w ,

$$Q_1(w) = Q_1(y_1; w), \text{ and } Q_{j+1}(w) = [y_{j+1}, Q_j(w) \setminus \{y_j\}], \text{ for } 1 \leq j \leq s-1.$$

Lemma 3. *For each j , $1 \leq j \leq s$, the algorithm invokes $Q(y_j, I; w)$ for some tuple I and each man in $\{x_{i_{j-1}+2}, \dots, x_{i_j}, y_{j+1}\}$ is active in it.*

Proof. By induction. The base case is given by y_1 , for which it has been shown earlier that $Q(y_1; w)$ is invoked at the beginning and every man in $\{x_2, \dots, x_{i_1}, y_2\}$ is active in $Q(y_1; w)$ after y_1 . Thus, the base case has been proved.

Suppose that the induction hypothesis holds for y_t , where $t \leq s-2$, *i.e.* for some tuple I , $Q(y_t, I; w)$ is invoked, and each man in $\{x_{i_{t-1}+2}, \dots, x_{i_t}, y_{t+1}\}$ is active in it. We will complete the proof by showing that the hypothesis holds for y_{t+1} . If y_{t+1} is “new”, *i.e.* it is added to N during the invocation of $Q(y_t, I; w)$, then $Q(y_{t+1}, y_t, I; w)$, is invoked subsequently. Using the fact that y_{t+1} is active in both $Q_{t+1}(w)$ and $Q(y_t, I; w)$, and all the men in $\{x_{i_t+2}, \dots, y_{t+2}\}$ are active in $Q_{t+1}(w)$ after y_{t+1} , 2 applied to each of them, implies that they are also active in $Q(y_{t+1}, y_t, I; w)$. Hence, for this case, the hypothesis is proved for y_{t+1} .

If y_{t+1} is already in N when $Q(y_t, I; w)$ is invoked, then for some tuple I' , y_{t+1} should have been added to N , during the invocation of $Q(I'; w)$. Thus, $Q(y_{t+1}, I'; w)$ would have been invoked prior to $Q(y_t, I; w)$. Using the same argument (on $Q_{t+1}(w)$ and $Q(y_{t+1}, I'; w)$) that we used for the earlier case, we conclude that even for this case, the hypothesis holds for y_{t+1} . \square

Thus, we have shown that all the men in $AS'(w)$ are active somewhere during the execution of the algorithm and thus, all are present in N at the end of the execution. This completes the proof of 1. \square

Algorithm 1: $\mathcal{A}(Q, w)$, [TST01]

Input: Strategy Q , and a woman $w \in W$

Output: Sets $N = \{m \in M \mid \exists Q' \in \mathcal{Q}_w, \mu_{Q'}[w] = m\}$, and

$L = \{Q(m, I; w) \mid m \in N, Q' = (Q(-w), Q(m, I; w)), \mu_{Q'}[w] = m\}$

- 1 Let x_1 be the first active man in $Q(w)$;
- 2 Let $N \leftarrow \{x_1\}$ and $L \leftarrow \{Q(x_1; w)\}$;
- 3 **Explore**($Q(x_1; w)$);

Procedure **Explore**($Q(x; w)$)

Let $A \leftarrow \{\text{men who are active in } Q(x; w) \text{ after } x\}$;

foreach $y \in A \setminus N$ **do**

$N \leftarrow N \cup \{y\}$ and $L \leftarrow L \cup \{Q(y, x; w)\}$;
Explore($Q(y, x; w)$);

return (N, L);

4 Algorithm for P -Stable Nash Equilibrium

In this section, we consider the problem of deciding, for a given true preference P , and stated preference Q , whether Q is P -stable Nash equilibrium. We show that this problem is solvable in time $O(n^4)$. Algorithm 2 uses Algorithm 1 as a subroutine.

To show the correctness of Algorithm 2, the following lemma plays a key role.

Algorithm 2: Algorithm for testing P -stable Nash equilibrium

Input: True strategy P , stated strategy Q , and the set of women W .

Output: Answers “Yes”, if Q is a P -stable Nash equilibrium, else “No”.

```
foreach  $w \in W$  do
    Run Algorithm 1 to obtain  $N_w(Q)$  and  $L_w(Q)$ ;
    Let  $N' \leftarrow \{m \in N_w(Q) \text{ s.t. } w \text{ prefers } m \text{ to } \mu_Q[w] \text{ w.r.t } P(w)\}$ ;
    foreach  $m \in N'$  do
        Let  $Q(m, I; w) \in L_w(Q)$  be the list that yields  $(m, w)$  as a matched pair;
        Compute  $\mu$ , the men-optimal stable matching for  $(Q(-w), Q(m, I; w))$ ;
        if  $\mu$  is  $P$ -stable then
            return “No”;
return “Yes”;
```

Lemma 4. Suppose that Q is an arbitrary (unilateral) manipulation strategy of woman w . Then for any (fixed) man \bar{m} , a Q -stable matching (if any) that matches \bar{m} to w , is unique.

Proof. Let Q be an arbitrary manipulation strategy by which w attains \bar{m} , and let μ denote the men-optimal stable matching for Q .

Algorithm 1 computes a list $Q(\bar{m}, I; w)$, such that w attains \bar{m} by the strategy $Q^* = (Q(-w), Q(\bar{m}, I; w))$. Note that \bar{m} appears at the front of the list $Q^*(w) = Q(\bar{m}, I; w)$. Let μ^* be the men-optimal stable matching for Q^* . Our goal is to show that $\mu = \mu^*$.

Claim 1. For each m , $\mu^*[m] \geq \mu[m]$, in $Q(m)$.

Proof. We begin by showing that μ is Q^* -stable. Note that μ is Q -stable, and Q and Q^* differ only in w 's list. Hence, if there is a Q^* -blocking pair in μ , then it must contain w . However, this is impossible since w is matched with \bar{m} , who is at the front of the list $Q^*(w)$. Therefore, μ must be Q^* -stable.

Since μ^* is the man-optimal stable matching for Q^* and μ is a Q^* -stable matching, consequently, for each man $m \in M$: $\mu^*[m] \geq \mu[m]$, in $Q^*(m)$. Since $Q^*(m) = Q(m)$, for each man m , the claim is proved. □

Claim 2. For each m , $\mu[m] \geq \mu^*[m]$, in $Q(m)$.

Proof. We will show that μ^* is Q -stable. Suppose that it is not. Then there is Q -blocking pair in μ^* , and it includes w for the same reason as in the proof of 1. Let (m', w) denote a Q -blocking pair. Then $w > \mu^*[m']$, in $Q(m')$, and $m' > \mu^*[w]$, in $Q(w)$.

By 1, m' prefers w to $\mu[m']$, and recall that $\mu^*[w] = \mu[w] = \bar{m}$. Hence, (m', w) is a Q -blocking pair in μ , a contradiction. Again, for the same reason as in the proof of 1, we can conclude that $\mu[m] \geq \mu^*[m]$, in $Q(m)$, for each man m . □

By Claims 1 and 2, $\mu[m] = \mu^*[m]$ for each man m . Thus, $\mu = \mu^*$, completing the proof of 4. □

Theorem 2. Algorithm 2 solves our Problem in $O(n^4)$ time.

Proof. For each woman, Algorithm 2 runs Algorithm 1 whose time complexity is $O(n^3)$. The size of set N is at most n , and Algorithm 2 runs GS-M, whose time complexity is $O(n^2)$, on $(Q(-w), Q(m; w))$ for each $m \in N$. Therefore, its running time is $O(n^3)$ for each woman. Since there are n women, the total running time of Algorithm 2 is $O(n^4)$.

Suppose that Algorithm 1 outputs “No”. Then, it implies that a P -stable manipulation strategy was found by Algorithm 1, and therefore Q is not P -stable Nash equilibrium. For the opposite direction, suppose that Q is not a P -stable Nash equilibrium and there exists a woman w who has a P -stable manipulation strategy Q' . Then $\mu_{Q'}[w]$ is added to N when Algorithm 1 is run for w , and by 4 the matching $\mu_{Q'}$ is uniquely defined. Since $\mu_{Q'}$ is P -stable, Algorithm 2 must output “No.” \square

References

- [DF81] L. E. Dubins and D. A. Freedman. Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.
- [DGS87] G. Demange, D. Gale, and M. Sotomayor. A further remark on the stable matching problem. *Discrete Applied Mathematics*, 16:217–222, 1987.
- [DST15] Y. Deng, W. Shen, and P. Tang. Coalition manipulations of the gale-shapley algorithm. *arXiv:1502.07823*, 2015.
- [Ehl08] L. Ehlers. Truncation strategies in matching markets. *Mathematics of Operations Research*, 33(2):327–335, 2008.
- [GI89] D. Gusfield and R. W. Irving. *The Stable Marriage Problem-Structure and Algorithm*. The MIT Press, 1989.
- [GS62] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.
- [GS85] D. Gale and M. Sotomayor. Ms. machiavelli and the gale-shapley algorithm. *American Mathematical Monthly*, 92(4):261–268, 1985.
- [Hua06] C-C. Huang. Cheating by men in the gale-shapley stable matching algorithm. In *ESA '06*, volume 4168 of *LNCS*, pages 418–431, 2006.
- [III⁺13] T. Inoshita, R.W. Irving, K. Iwama, S. Miyazaki, and T. Nagase. Improving man-optimal stable matchings by minimum changes to preference lists. *Algorithms*, 6(2):371–382, 2013.
- [IM05] N. Immorlica and M. Mahdian. Marriage, honesty and stability. In *Proceedings of SODA '05*, pages 53–62, 2005.
- [KM10] H. Kobayashi and T. Matsui. Cheating strategies for the gale-shapley algorithm with complete preference lists. *Algorithmica*, 58:151–169, 2010.
- [Man13] D. F. Manlove. *Algorithmics of matching under preferences*, volume 2 of *Theoretical Computer Science*. World Scientific, 2013.
- [PRVW11] M. S. Piny, F. Rossi, K. B. Veneble, and T. Walsh. Manipulation complexity and gender neutrality in in stable marriage procedures. *Auton. Agent Multi-Agent Systems*, 22(183–199), 2011.

- [RR99] A. E. Roth and U. G. Rothblum. Truncation strategies in matching markets-in search of advice for participants. *Econometrica*, 67(1):21–43, 1999.
- [RS90] A. E. Roth and M. Sotomayor. *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*. Cambridge Univ. Press, 1990.
- [TST01] C-P. Teo, J. Sethuraman, and W-P. Tan. Gale-shapley stable marriage problem revisited : Strategic issues and applications. *Management Science*, 47(9):1252–1267, 2001.